2016

Plant Localization and Discrimination using 2D+3D Computer Vision for Robotic Intra-row Weed Control

Jingyao Gai
Iowa State University, jygai@iastate.edu

Lie Tang
Iowa State University, lietang@iastate.edu

Brian Steward
Iowa State University, bsteward@iastate.edu

Plant Localization and Discrimination using 2D+3D Computer Vision for Robotic Intra-row Weed Control

Abstract
Weed management is vitally important in crop production systems. However, conventional herbicide based weed control can lead to negative environmental impacts. Manual weed control is laborious and impractical for large scale production. Robotic weed control offers a possibility of controlling weeds precisely, particularly for weeds growing near or within crop rows. A computer vision system was developed based on Kinect V2 sensor, using the fusion of two-dimensional textural data and three-dimensional spatial data to recognize and localized crop plants different growth stages. Images were acquired of different plant species such as broccoli, lettuce and corn at different growth stages. A database system was developed to organize these images. Several feature extraction algorithms were developed which addressed the problems of canopy occlusion and damaged leaves. With our proposed algorithms, different features were extracted and used to train plant and background classifiers. Finally, the efficiency and accuracy of the proposed classification methods were demonstrated and validated by experiments.

Keywords
Robotic weeding, Computer vision, Sensor fusion

Disciplines
Agriculture | Bioresource and Agricultural Engineering

Comments
This paper is from 2016 ASABE Annual International Meeting, Paper No. 162460814, pages 1-15 (doi: 10.13031/aim.20162460814). St. Joseph, Mich.: ASABE.. Posted with permission.

2950 Niles Road, St. Joseph, MI 49085-9659, USA
269.429.0300 fax 269.429.3852 hq@asabe.org www.asabe.org

An ASABE Meeting Presentation
***DOI:** 10.13031/aim.20162460814*
***Paper Number:** 162460814*

Plant Localization and Discrimination using 2D+3D Computer Vision for Robotic Intra-row Weed Control

Gai, Jingyao1; Tang, Lie1; Steward, Brian1
[1]Iowa State University

Written for presentation at the
2016 ASABE Annual International Meeting
Sponsored by ASABE
Orlando, Florida
July 17-20, 2016

ABSTRACT. *Weed management is vitally important in crop production systems. However, conventional herbicide based weed control can lead to negative environmental impacts. Manual weed control is laborious and impractical for large scale production. Robotic weed control offers a possibility of controlling weeds precisely, particularly for weeds growing near or within crop rows. A computer vision system was developed based on Kinect V2 sensor, using the fusion of two-dimensional textural data and three-dimensional spatial data to recognize and localized crop plants different growth stages. Images were acquired of different plant species such as broccoli, lettuce and corn at different growth stages. A database system was developed to organize these images. Several feature extraction algorithms were developed which addressed the problems of canopy occlusion and damaged leaves. With our proposed algorithms, different features were extracted and used to train plant and background classifiers. Finally, the efficiency and accuracy of the proposed classification methods were demonstrated and validated by experiments.*

Keywords. *Robotic weeding, computer vision, sensor fusion*

Instructions

In recent years, with the development of greater health consciousness, consumers are becoming more and more interested in vegetables, especially natural, organic vegetables. In 2013, U.S. vegetable production resulted in a revenue of 11.4 billion dollars for fresh products, up 14 percent from 2012, with a total harvested area of 1.63 million acres (USDA-NASS, 2014). In organic vegetable production, USDA-NASS, (2014) stated there were 1.2 billion dollars in total for fresh productions, with total harvested area of 164,000 acres.

One important factor that affects crop yield is weed competition. Weeds are very competitive in obtaining moisture, sunlight, and nutrients, all of which which adversely affect crop yield and quality. The National Organic Farmers' Survey conducted by Walz (2004) reported that organic farmers indicated weeds were one of the major causes of production losses, second only to weather-related losses. In addition, weed management is also one of the most costly operations in vegetable production, especially for organic farming. Organic farmers have major production costs associated with weed control, which are mainly caused by the reliance on manual weeding. Earthbound Farms, the largest producer of organic vegetables in North America, reported that some of their farmers spent up to $1,000 per acre to control weeds manually (Zimdahl, 2013). It is obvious that labor costs have made manual weed control impractical. For some farmers, moreover, the use of herbicides is becoming less desirable due to the emergence of herbicide-resistant weeds, the environmental impact of herbicide runoff, as well as societal demand for chemical-free foods (McErlich & Boydston, 2014). Realizing that, many tools have been developed to increase mechanical weeding efficacy in recent years.

Inter-row (between crop rows) mechanical weeding is relatively easy to achieve by using commercial tools. However, intra-row (within or close to crop rows) mechanical weeding has a risk of damaging crop plants. To data, there are limited tools for intra-row weeding in vegetable crops with desirable weed control efficacy. With the advancement of computational technology, automated robotic weeding, especially with computer vision, offers a possibility of controlling weeds in a precise fashion, particularly for weeds growing near crops or within crop rows. Since identification and localization of plants have not yet been fully automated, research to address these problems is thus in great demand.

Related work

Research work has been reported in the literature for robotic weeding based on computer vision. With sensing method as criteria, most of those computer vision systems can be categorized into two classes: tow-dimensional (2-D) vision systems and three-dimensional (3-D) vision systems.

2D vision applications:

A 2D sensor is a type of sensor that can record light or other electromagnetic radiation reflected or emitted from objects, by focusing it on a light-sensitive surface. The word “2D” means horizontal and vertical dimensions in the image space, as it is a projection of the 3D real world.

Regular 2D cameras were the earliest ones used in robotic weeding. The problem of plant identification was addressed by extracting color or morphological features from a leaf or a whole plant, such as length, width, perimeter dimensions, roundness, circularity, convexity and moment. Slaughter, (2008) reviewed these types of systems and concluded that they demonstrated high recognition accuracy only under ideal conditions, in which light was controlled and plants were sparse in images. Also, they are not robust to occlusion problems, or defects of the plants caused by insect damage or wind, which are common in field. Moreover, with complex illuminant conditions, such as strong sunlight, images will be saturated. Most of the algorithms will fail to segment plants out of the background.

Spectral reflectance characteristics of plants were reported to be effective in vegetation segmentation and crop/weed discrimination (Scotford & Miller, 2005; Zwiggelaar, 1998). Zwiggelaar, (1998) also reviewed that the selected spectral wavebands for classification are generally different for different weed and crop pairs. Therefore selecting wavebands and designing algorithms for distinguishing crop plants from different weed species is complex. On the other hand, an NDVI (normalized difference vegetation index) map, which can be generated from an image from an infrared camera or NDVI sensor, is reported effective in vegetation segmentation. (Sui, Thomasson, Hanks, & Wooten, 2008)have developed a vision-based system for weed mapping using an NDVI camera. However, in order to accomplish crop/weed discrimination with NDVI images, morphological features are still needed. The advantages of spectral reflectance based methods are: they are less sensitive to environmental light, and the infrared reflectance of plants can include additional information for plant discrimination.

3D sensing applications:

A 3D sensor is a type of sensor that can measure the distances between objects and sensors, which are “depth”. The word “3D” corresponds to the three dimensions of the real world. In recent research, with the development of sensing technology, a 3D sensor promises to address problems in 2D vision systems, such as occlusion. A 3D sensor can also give reliable information to perform plant discrimination and plant localization. In several studies (J. Li, 2014; Jin & Tang, 2009; Nakarmi & Tang, 2012), 3D sensors were applied in agricultural applications which provided a good performance. The advantages of 3D sensors for plant discrimination and localization are obvious: 3D sensors can provide fundamental depth information, making it is much easier to obtain the 3D structural and morphological data of the plants.

Today, three types of state-of-art 3D sensors are mainly used on mobile robots: stereo vision, laser, and PMD time-of-flight. In several articles (Sansoni, Trebeschi, & Docchio, 2009; Weiss & Biber, 2011), the authors have compared and evaluated those three types of sensor for mobile robot:

To receive 3D data using stereo vision, typically triangulation of two cameras or structure-from-motion technique are used. In the study by Jin & Tang, (2009), a real-time sensing system for corn phenotyping was developed based on a stereo vision sensor. However, due to the stereo camera’s passive operation mode, it is hard to provide reliable data for accurate sensing: When receiving 3D data from stereo vision, the disparity calculation highly depends on the structures or features of the objects in images. Further, the precision and maximum depth are limited by the baseline between the cameras. Also, qualities of distance values decrease very fast as depth increases. The advantage of stereo vision is its high image resolution, with color information. One the other hand, cameras in stereo vision systems can also be modified into near-infrared cameras by replacing their filters with NIR filters, in order to take advantage of spectral reflectance information (Hunt et al., 2010).

The accuracy, resolution, frame rate as well as price of different 3D laser sensors are widespread. Some of the 3D laser sensors are implemented from 2D laser scanners which uses line-scanning method, such as Lidar. One example is Kurt3D (Surmann & Nüchter, 2003) which equips a rotating 2D SICK® laser sensor to realize 3D laser scanning. However, it needs a stop-and-go mode for traveling to receive consistent 3D data, since a 2D line sensor was not built for 3D applications. Generally, 3D laser sensors usually have the properties of high weight, high power consumption, as well as high price. Those make those 3D sensors less desirable for small mobile robots.

The semiconductor based PMD time-of-flight (TOF) camera is the latest technique. It measures distance and infrared reflectance intensity information based on time-of-flight technique. A modulated light signal is emitted by sensor, reflected by objects then received by sensor receiver (L. Li, 2014). The distances are calculated by the phase shift of the signal as well as the reflection intensity. There are many research applications using TOF cameras as sensing devices in agriculture, such as phenotyping (Alenyà, Dellen, Foix, & Torras, 2012), and plant spacing (Jin & Tang, 2009). In study by Kazmi, Foix, Alenyà, & Andersen, (2014), the authors summarized the advantages along with drawbacks of TOF cameras. The advantages are: they deliver high frame rates as well as accurate depth data under suitable conditions. The limitations are: resolutions of depth images are often low; the sensors are sensitive to ambient sunlight, which usually leads to poor performance while working outdoors; the quality of depth values depends on color of objects; and some sensors have blurring problems while sensing moving objects.

Although there are limitations for 3D cameras, 3D sensing is still beneficial in agricultural applications. For indoor applications like phenotyping facilities, very accurate depth measurements of plant organs are required to rebuild fine 3D models of plants; In field operations like weed control, 3D information can help not only in improvement of plant recognition and localization (J. Li, 2014) by resolving problems of occlusion, but also in estimation of infection, in order to apply precise amount of chemicals onto the crops (Nielsen, Andersen, Slaughter, & Giles, 2004).

It is clear that 2D images have higher resolution, more detail, and 3D images contain spatial information of plants. However, the prior works (cited above) paid considerably less attention to fuse different type sensors to increase the capacity. As a result, the novel contribution of this study over the existing studies lies in:

- Fuse two-dimensional textural data and three-dimensional spatial data to accomplish plant localization and discrimination, which takes advantages from both sensor types.
- Develop a new image processing method accomplishing crop plant detection and localization accurately against different weed species at different growth stages

Materials and methods

Sensor

Kinect v2 sensor (Figure 2-1(a)) was used in this study. Kinect v2, developed by Microsoft, provides color (RGB), IR (Infrared) and depth information. The 3D sensor in Kinect v2 is a semiconductor based PMD (photon mixer devices) sensor based on Time-of-Flight principle. The 2D color information can be registered into the 3D space point with sensors' relationship provided in Kinect SDK (Figure 2-1(b)). The specification of Kinect v2 is listed in Table 2-1. Kinect v2 sensor uses three strong IR emitters as light sources (Figure 2-2). They enable the Kinect v2 to work outdoors under indirect sunlight, though it is still not operational under direct sunlight.

Figure 1. (a) Kinect v2 sensor used in this study. (b) The colored point cloud output from fusing 3D depth information and 2D color information.

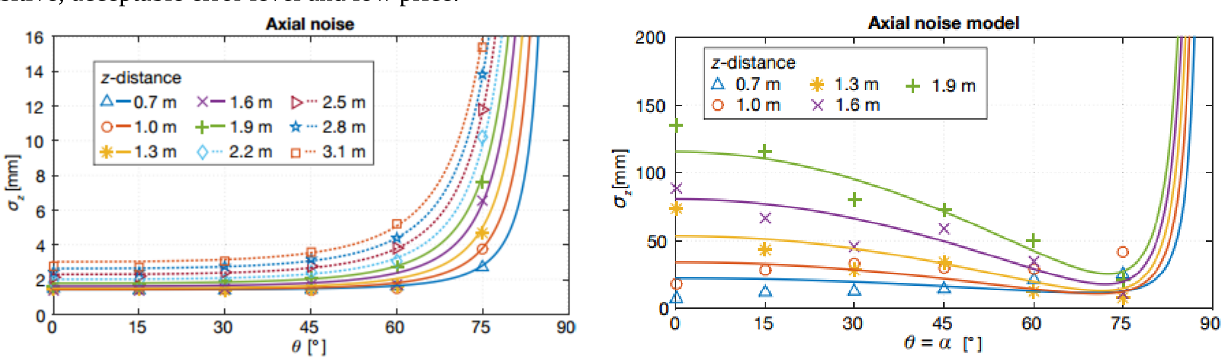
Table 1. Kinect v2 Specifications

Infrared/depth camera	Resolution	512 x 424 px
	Field of view (h x v)	70º x 60º
	Operating range	0.5 – 4.5 m
	Depth resolution	1 mm
Color camera	Resolution	1920 x 1080 px
	Field of view (h x v)	84º x 54º
	Color depth	256 bit, 3 channels
Frame rate		30 Hz
Shutter type		Global shutter
Voltage		12 VDC
Power usage		15 W
Dimensions(w x d x h) (mm)		249 x 66 x 67
Mass		970 g
Price		199 USD

Accuracy is an important parameter to evaluate a sensor in this project. The error is composed of both non-systematic and systematic error. The non-systematic error is characterized by statistical uncertainty or noise level. A comprehensive report of evaluating the noise level of Kinect v2 was reported by Fankhauser et al., (2015). The author tested the axial noise σ_z (noise level along the z-axis) of the depth output at different distances and angles of the observed surface. Both indoor and outdoor situations are tested. The results are shown in Figure 2-3. The author also found that the noise levels are similar when testing indoor and testing outdoors with overcast sky conditions, and the noise level is significant when testing in direct sunlight. For this research project, the distance between the sensor and objects is within 0.75-1.25 m, the noise level σ_z should be around 2 mm. Thus, it is clear that the sensor is competent to output reliable data under indirect sunlight.

Equally important, systematic error, which is a type of error that deviates by a fixed amount from the true value of measurement, is also evaluated by Fankhauser et al., (2015). The author analyzed the systematic errors such as depth distortion, amplitude-related error, and temperature-related error. The composited systematic error is still in millimeter level when working at a short working distance and perpendicular to the object surface. But it increases with working distance, or angle of the object surface. In addition, Corti, et al. (2015) also found that different materials, surfaces, as well as different colors, will result in small offset in depth measurement. Still, it is not significant, which was estimated to be ±1mm.

Overall, the Kinect v2 sensor was found competent for this project, as a result of its high resolution, illuminant insensitive, acceptable error level and low price.



Axial noise level in indoor test

Axial noise level in outdoor test

Figure 2. Axial noise level (standard deviation of measured depth values) at different working distances (z-distance) and different angles of observed object surface (θ), in both indoor (a) and outdoor (b) tests (Fankhauser et al., 2015). The θ is defined as the angle difference between the direction of sensor and the surface normal.

Location and agronomic trial

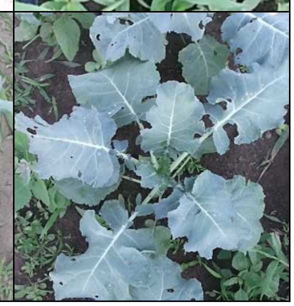
The target crop species on which algorithms were developed and tested in this study are: lettuce (Lactuca, L.), broccoli (Brassica oleracea L. var. botrytis L.). The negative input of the system are various types of weeds that are common in Iowa, including bromegrass (Bromus inermis Leyss), pigweed (Amaranthus spp.), Lambsquaters (Chenopodium album), waterhemp (Amaranthus rudis), barnyardgrass (Echinochloa crus-galli), bindweed (Convolvulus arvensis), purslane (Portulaca oleracea), and white clover (Trifolium repens).

The data used in this study were obtained in the horticulture research station of the Iowa State University in Story County (42.11 º N, -93.59 º E). The soil is nearly level to moderately sloping landscape near the Skunk River. Soil type was Clarion loam, moderately eroded, with 5 to 9 percent slope. The average annual temperature is 49.45ºF and the average annual precipitation is 35.83 inch.

The field reserved for acquisition in this study was 1/8 acres. The crop plants were started in greenhouse and transplanted in the horticulture research station with row spacing of 30 in and plant spacing within the same row of about 12 in. Weeding on the field was not performed within 5 days before data acquisition, in order to make sure crop plants and weeds are observed in the same images.

The image data was collected with Kinect v2 sensor mounted on a pushing cart in afternoons on both sunny and cloudy days. Data was acquired 4 time for each species with 5 days interval on average. In sunny days, an umbrella was used to block the sunlight to increase the quality of the images. In this project, the data collection cart was pushed at the speed of about 0.3m/s between crop rows. 3D depth images, as well as 2D color images, were taken about every two seconds. In total, 1156 images were taken from 272 plants at four different times.

Table 2. Examples of 2D images taken using Kinect v2 sensor from different species in different collected time

Species	Stage 1	Stage 2	Stage 3	Stage 4
Lettuce				
Broccoli				

Algorithm design

In order to accomplish plants detection and localization, 3D spatial and 2D color information are fused together. The algorithm was developed using OPENCV library, C++ (Bradski & Kaehler, 2008) and R. The framework of the algorithm is shown in the flowchart in Figure 2-6 and stated as follows:

The algorithm takes a depth image which contains 3D spatial information with corresponding 2D color information acquired through a Kinect V2 sensor as input.

Step 1: Preprocessing. It is needed to remove roughly the invalid pixels and remove noise points in point clouds. This procedure is done by using a usable-are filter, a cut-off filter and a simplified neighbor count filter.

Step 2: Registration and segmentation. 2D and 3D information are fused. The ground is detected using both 2D color and 3D depth information, and a plane equation is fitted using RANSAC. All vegetation pixels are extracted in the images.

Step 3: Clustering. The remaining points which belong to plants are separated into different clusters, in which each cluster corresponds to a candidate of a crop plant. Each cluster may contain one plant, or more plants which are connected in space, with species not determined.

Step 4: Feature extraction and grouping. Morphological and structural features in each cluster are extracted. Those features also used to find out how many crop plant candidates are included in each cluster.

Step 5: Classification and localization. Classification is applied to all the plant candidates based on their features, to determine whether they are crop plants or not. The positions of the plants are determined according to the fitted ground plane and their features as well, such as veins and canopy canters.

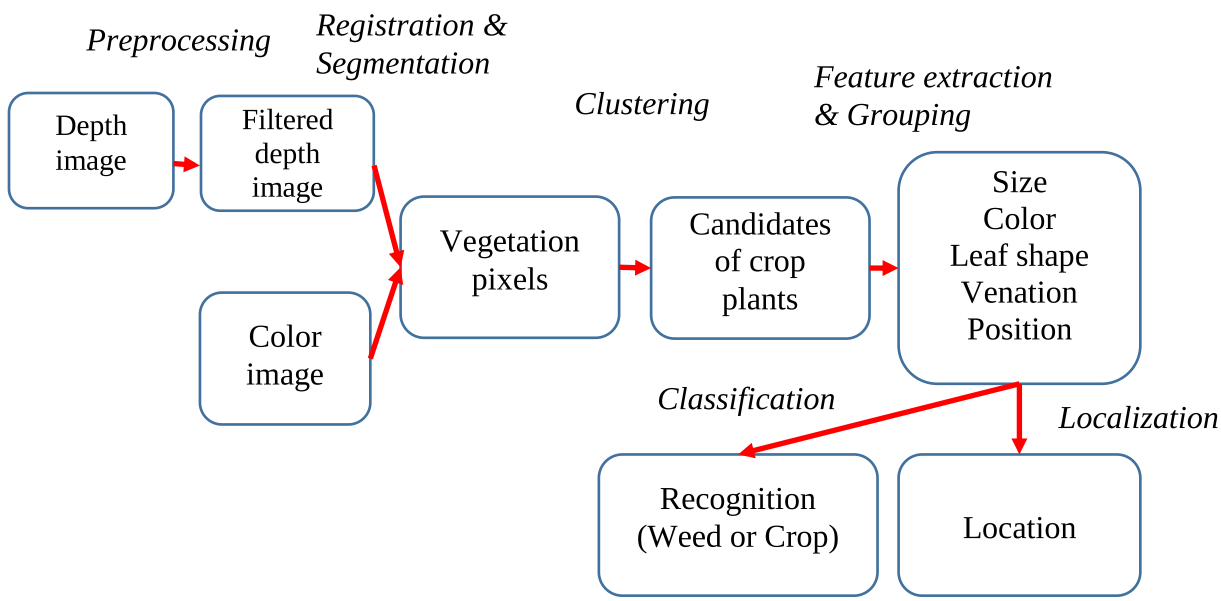


Figure 3. The data flowchart of the image processing system

Preprocessing on depth image

In general, data collected by optical sensors contains noise, which will cause unexpected results without filtering those noise. This is also true for Kinect V2 depth sensor, especially when working outdoors. Within preprocessing, useful data (in contrast to sparse noise & sensor bad points) is extracted from the depth image, against the disturbances from noise and useless information (such as data collection cart wheel). In our algorithm, three simple filters are applied on the depth image: a useable-area filter, a depth cut-off filter, and a neighbor count filter.

The first filter to apply is a reliable area filter. Because of the ambient light, the off-center pixels in depth images of Kinect V2 sensor have a higher chance of carrying incorrect depth information. In this study, a round area with a radius of 220 pixels was selected as reliable area by observing and testing. In this process, about 30% of the points are deleted.

The second filter to apply is a depth cut-off filter, which is used to remove pixels laying outside of a defined depth range. With this filter, bad points from the sensor (usually have a distance of zero or infinity) and points with significant depth value (noise in most cases) are removed.

At last, a simplified neighbor count filter was modified from Statistical Outlier Filter developed by (Rusu, 2009) to remove sparse noise. For depth sensors such as Kinect V2, the pixels are denser in real world on flat surfaces than on uneven surfaces. In another word, the pixel on a flat surface has more connected pixels in 3D space. Those “connected” pixels are called “neighborhood” in 3D point cloud. The concept of “connectivity” in 2D images can be applied in this case. The potential neighbors can only be the surrounding pixels on depth image. A threshold of 10 mm was defined as the limit of the distance of deciding neighborhood based on sensor error and the spatial resolution. If less than 15 neighbors exists for an individual testing point in a 5x5 search window, it is considered as a sparse noise.

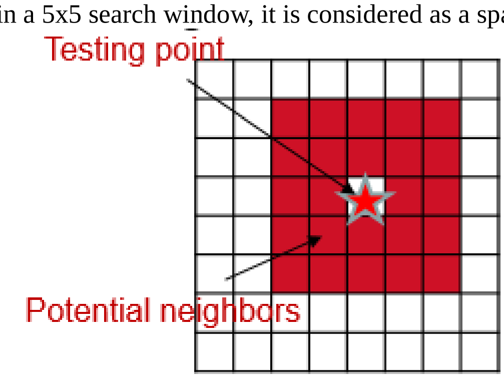


Figure 4. The strategy of the neighbor counts filter.

Segmentation using depth and color

In segmentation, the pixels are divided into two different subsets, one is the vegetation pixels set, and the other is the background pixels set. Both depth and color information are utilized in this procedure. It is obvious that vegetation pixels are higher in green color and higher in depth, while background pixels are darker in color and lower in depth.

Color based segmentation:

In color based segmentation, the main strategy is to find the green pixels. However, segmentation using RGB channels only is unstable. The complexity of illumination conditions such as shadows will challenge the color-based segmentation algorithms. However, illuminant invariant space and HSV color space are found useful in extracting green pixels, which are less sensitive to illumination conditions.

Finlayson et al. (2002) proposed a method using illuminant invariant (ill-inv) space to remove the shadows. It transfers the color of each pixel from RGB color space into ill-inv color space, where the Y axis is $\log(B/R)$, and the X axis is $\log(G/R)$. In ill-inv space, the same color under different light conditions forms a line, and all lines of different colors are approximately parallel. If project those ill-inv color space points onto a specific line, which is called invariant axis, all different colors can be separable.

An illuminant-invariant image I_{il} can be generated and saved to assistant the segmentation procedure. One example of the illuminant invariant image is shown in Figure 5(b), which reduced the effect of shadow compared with the original color image Figure 5(a).

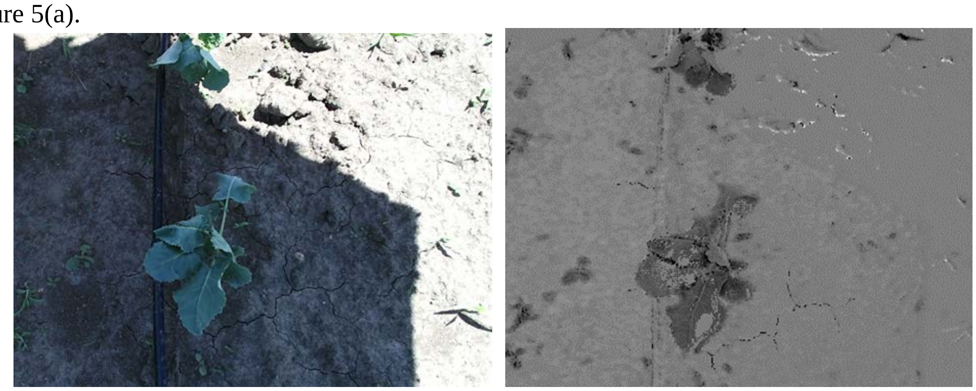


Figure 5: The effect of illuminant-invariant map. In the left figure, the light conditions are different between shaded and unshaded area. And in the right figure, the illuminant-invariant map is generated and compensates the effect of changing light conditions. Green pixels have lower values in illuminant-invariant maps.

In the research of Philipp & Rath, (2002), the author also found the HSV (Hue-Saturation-Value) color space to be one of the most reliable color space to distinguish green plants from the background. The hue values (H channel) of plants don't change significantly with different light intensity. Thus, the HSV color space image I_{HSV} is also stored for the next segmentation step.

Depth based segmentation:

In depth based segmentation, the strategy of the algorithm is to identify the ground, which can be assumed as a plane. The ground identification is based on depth information, as well as the color information extracted in the previous step. The ground is visible in images because the sensor is looking down at the plants. Ideally, the individual plants could be separated after fitting the ground plane and eliminating the ground pixels.

In order to fit the ground plane, the weighted Random Sample Consensus (RANSAC) algorithm, as well as the robust regression are used. RANSAC is used as a rough ground fitting method, and the robust regression is used for refining the ground plane model.

Since pixels belonging to the ground should be close to the fitted ground plane, they should have higher weights while performing fitting. In contrast, the plant pixels are more likely to be the outliers when fitting the ground, then they should have lower weights. One solution is to define the weight function to be:

$$Weight(p) = 1 + P(p \in Ground) \tag{1}$$

In which p is one pixel in the point cloud, and the P(p∈Ground) means the probability of p is belonging to the ground pixel set.

Based on Bayes' law and data collected, the weight function for working with lettuce and broccoli are decided using logistic function. In this study, considering the computational cost of the program, the variable is selected from values in three color spaces: RGB, HSV, or Illuminant variant. After fitting models and evaluating them using ROC curve, the models are selected as:

Lettuce:

$$Weight(p) = 1 + \frac{1}{1 + \exp(2.444 + 5.831 * ill)} \tag{2}$$

Broccoli:

$$Weight(p) = 1 + \frac{1}{1 + \exp(3.438 + 0.156 * R - 0.073 * G - 0.087 * B)} \quad (3)$$

Where R, G, B are color values in RGB space (0-255) and ill is value in illuminant-invariant space.

Plant clustering

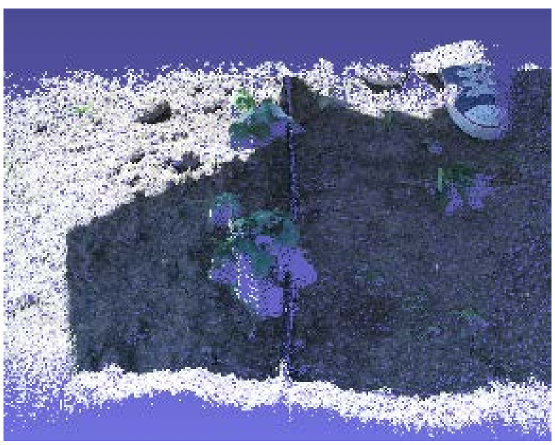
Plant clustering is performed to assign the remaining points into different clusters representing different plant candidates. This clustering problem can be described as an unsupervised 3D clustering problem. It is unsupervised because the spatial continuity is the only information can be relied on in this clustering problem. Different methods have been implemented (except for Superparamagnetic clustering (Blatt, Wiseman, & Domany, 1996)) and applied to solve this clustering problem. A brief comparison is in Table 3.

After considering both advantages and drawbacks, the 2D connected components method was finally adopted, because of its high efficiency, and its negligible drawbacks. Due to the fact that the data's range in depth (or z-axis) is very narrow compared with the other two dimensions (x, and y-axis), thus it is difficult for any clustering method to rely on depth information to separate individual plants perfectly. Further analysis will be performed after extracting features.

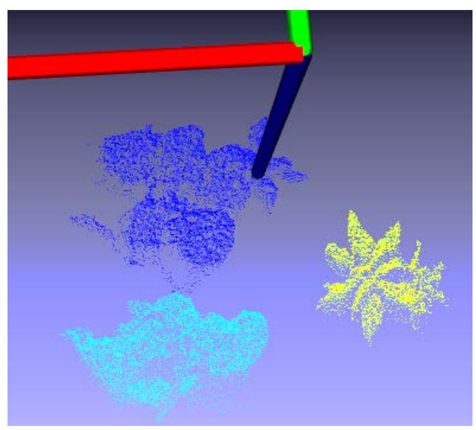
After clustering, small clusters containing fewer pixels than a predefined threshold are no longer kept, since they can only be weeds or noise. Thus, the plant data point set is divided into different subsets representing different plant candidates.

Table 3. Table showing the description of different clustering methods, as well as their advantages and disadvantages.

Name	Description	Pros	Cons
K-Mean	Iteratively find K clusters to minimize the sum of with-in cluster variation.	High efficiency	K need to be specified
ISODATA	Modified K-Mean for unknown number of clusters	High efficiency, Algorithm splits and merges clusters	Unexpected results when plants are connected
Superparamagnetic clustering	Animation of the ferromagnet movement	High quality results	Not open source, hard to implement
Region Growing	Growing from seeds to neighbors, until the whole cloud is processed	High quality results	Low efficiency, since hard for compiler to optimize for parallelization
Connected components in 2D	Find connected components in depth image	Best efficiency	Change in depth (z-axis) not considered when clustering



(a) Point cloud of broccoli crops in field



(b) Clustering result

Figure 6. A run-time example showing the plants clustering procedure. Fig (a) is the point cloud generated from Kinect v2 sensor after preprocessing. And Fig (b) is the result after background removal and clustering. Different clusters are assigned with different colors. Ideally, each cluster stands for a plant.

Feature extraction and grouping

The features extracted mainly include morphological features, as well as some color features. The morphological features consist of features which can describe dimensions and shapes of canopies or leaves, such as plant height, diameter, leaf aspect ratio and leaf area convex hull ratio, etc. Color features include average hue and saturation, as well as illuminant-invariance to take use of the color difference between crops and weeds if possible. With those features, connected plants within same crop candidate clusters can be separated, and classifiers can be built for different species at different growth stages.

Here are the definitions of these features to be extracted, and some of them are referred to the work of (Du, Wang, & Zhang, 2007).

- Leaf morphological features:
  (1) Leaf Average Height (havg): The average height of a leaf is defined as the distance between the field ground and the center point of the leaf.
  (2) Leaf Area (area): The area of a leaf is defined as the area of the leaf projection to the ground plane.
  (3) Leaf Length (leaf_l): The leaf length is defined as the maximum distance on the leaf to the center of the plant.
  (4) Leaf Width (leaf_w): The leaf width is defined as the maximum width on the leaf, which is perpendicular to the length direction.
  (5) Leaf Aspect Ratio (ratio): The aspect ratio of a leaf is defined as the ratio between L_{max}: the maximum length of the leaf measured from the center of the plant, and W_{max}: the maximum width on the leaf
  (6) Leaf Roundness (leaf_roundness): The roundness of a leaf is defined as the ratio between the leaf contour length C_{leaf} and the bounding ellipse circumference C_e.
  (7) Leaf Rectangularity (leaf_recti): The rectangularity of a leaf is defined as the ratio between the leaf area A_{leaf} and the leaf's bounding rectangle area A_c.
- Leaf color features:
  (8) (9) Hue, Saturation (clr_hue, clr_setra): The hue and saturation of a leaf are characterized using the average values of Hue and Saturation channel in HSV (Hue-Saturation-Value) color space respectively.
  (10) Illuminant-invariant Value (clr_ill):The value in illuminant-invariant space
- Leaf structural feature:
  (11) Venations: The veins with arcuate, pinnate, or palmate pattern, which contain information of leaf directions.
- Canopy morphological features:
  (11) Height (z_height): The height of a plant is defined as the maximum perpendicular distance between the field ground and the highest point of the plant.
  (12) Radius (radi): The radius of a plant is defined as the average length of the long leaves in all directions. The short new grown leaves which are shorter than their underlying leaves are not considered.
  (13) Leaf Number (leaf_n): The leaf number is defined as the number of leaves that can be segmented from the plant data.
- Canopy color features:
  (14) (15) Hue, Saturation (clr_hue, clr_setra): The hue and saturation of a plant are characterized using the average values of Hue and Saturation channel in HSV (Hue-Saturation-Value) color space respectively.
  (16) Illuminant-invariant Value (clr_ill): The average illuminant-invariant value of the plant data points.

Leaf segmentation and grouping:

To acomplish plant localization and recognition, it is very important to segment each leaf out of the biomass pixels set. Most of the reliable features to be used for classification are leaf features, such as leaf length, width, shape and color. Furthermore, the topology of leaves is used for solving the occlusion and connection problems in further steps.

The main strategy is to use marker-controlled watershed segmentation algorithm on the depth image to separate leaves. Color information (Hue channel in HSV color space) is also used to assist the segmentation process as well.

In order to solve the connection problem, a method was developed for this study. Based on the sensor view field, the maximum number of crops can be observed in the same image is three. So three circles are created to represent three potential combinations of the leaves. Those circles are initialized at different positions at first. Then the circles move to the new positions by calculating the mean positions of pixels from their surrounding leaves iteratively, until their positions are converged (Figure 7). Therefore, three potential combinations of leaves are generated, and passed to the next step for further processing.

Localization:

For localization, if the plants are separable during the clustering producer, the geometrical center of the cluster will be used as the localization result. However, it is also very common that there are more than one plants connected and are in the same cluster. Additional information is needed to address such connection problems.

In this study, venation was found to be the most stable leaf morphological feature in localizing broad leaf crops such as broccoli. The veins grow linearly from the centers of the plants to each leaf, which describes the direction of leaf best. In addition, the venation close to the center is obvious and is not likely to be occluded by other plants, or damaged by small animals.

In this section, a new developed venation-based plant localization algorithm is proposed, which works on species with border leaves and obvious straight veins. The algorithm has two steps, the first is venation extraction, in which extract the vein pixels and convert them into line segments. The second is center finding using those line segments.

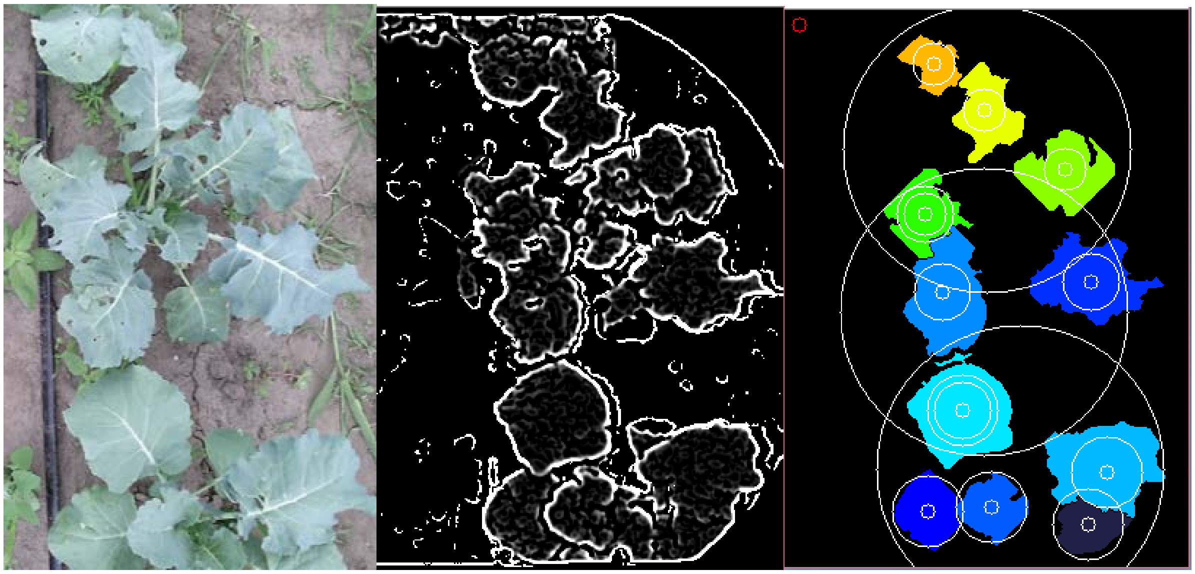


Figure 7. A run-time output example showing the leaf extraction procedures. Left: Color image of on cluster after segmentation and clustering, with connected crops, occlusions, as well as broken leaves. Middle: Combined gradient magnitude image on both color and depth. Right: Leaves extraction result example. Most of the leaves are extracted and labeled with different colors.

Veins are obvious in color images, especially in the Red channel from RGB color space images. As the pixels from the veins have higher values in red compare to surrounded green pixels, those vein pixels are considered as “ridges” in the images. In images, a “ridge” point has a lower negative eigenvalue in Hessian matrix, and a “valley” point has a higher positive eigenvalue in Hessian matrix. It can be seen that the veins are obvious on leaves in a map with the magnitude of negative eigenvalues of Hessian matrices for each pixel (Figure 8).

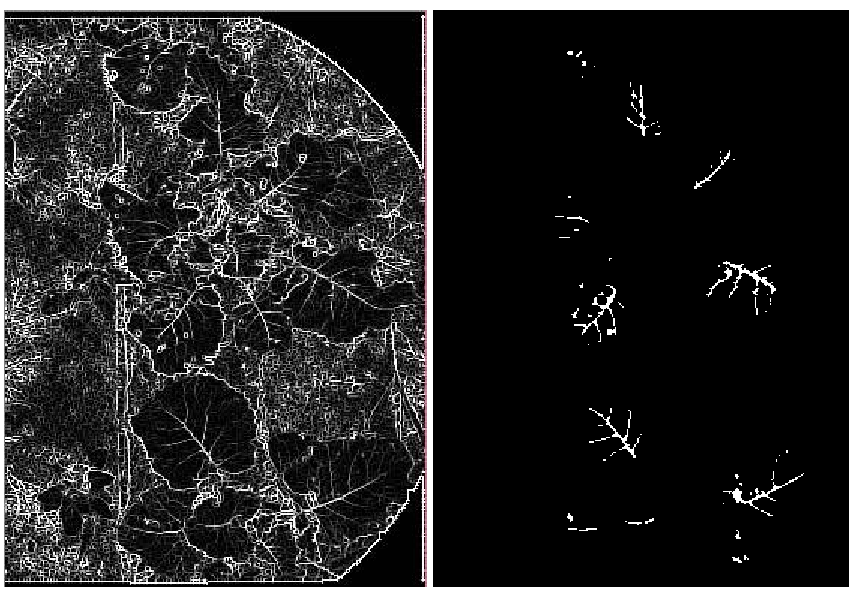


Figure 8. A run-time output example showing venation extraction methods. The left figure shows the map of lower eigenvalues of Hessian Matrixes at each location. Ridges are colored with brighter colors. The middle figure shows the veins extracted by thresholding. The right figures show the clean result after dilating, skeletonizing, and removing smaller clusters.

Suppose the primary veins’ direction can be extracted perfectly, then all the primary veins can be estimated as a line segment. Then, ideally, the center point lies on every extension lines of each line segment. If there are only two lines unparalleled, the point of intersection is unique. If there are more than three lines, the problem becomes finding the point, which is closest to all those extension lines. Weighted robust least square algorithm to solve the center finding problem.

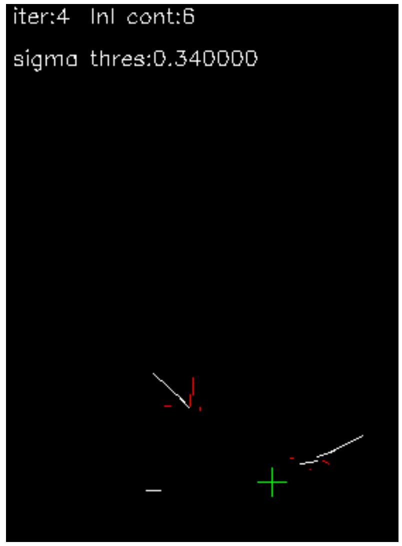


Figure 9. A run-time output example showing the iterations of finding the center. The green cross indicates the resultant center location in the current iteration. The line segments with white color are the active line segments (inliers, with non-zero weights). And the line segments in red are the inactive line segments (outliers, with zero weights) in the current iteration.

Classification

The purpose of classification is to predict the class of a new unknown sample point based on database that has been separated into different classes. All numeric features can be combined to feature vectors in order to build classifiers. As the leaf features can be extracted from plants of some species, such as broccoli, two classifiers are built. One is for leaves, and another is for plants. In leaves classification, the result should be whether those leaves are from crops or weeds, or they are incorrectly segmented. In the plant classification, the classifier uses the features of classified crop leaves, as well as the canopy features, then output whether the plant is a crop, or a non-crop.

There are many classification algorithms available, for instance the logistic regression (LR), the k-nearest neighbors (KNN), the artificial neural network (ANN), the Bayes classifiers such as LDA and QDA, the support vector machine (SVM), and tree-based classifiers such as decision tree and random forest. The performance of each classifier highly relies on the distributions of data. Thus, it is not suitable to subjectively decide which model is the best. In this study, PCA (principle component analysis) is performed first to ensure all the predictors are non-correlated, and all the above models are fitted using the features extracted with the method introduced. Then those models are evaluated and compared by using Cross-Validation (CV) results. All of the classification programs are implemented using R language.

Experimental results and discussion

Crop plant candidate detection

After testing the segmentation algorithm and clustering on the data collected, we found that about 95% of the crop plants can be segmented and detected as candidates correctly, in which the target crop plants are included in a candidate cluster. The reasons of the 5% failures are:

1. Sometimes, the weeds covered most of the area in the field of view. As the ground segmentation relies on the depth information of the ground pixels, it will cause incorrect results if the ground is not observable.
2. Sometimes, especially at early growth stages, the crops are lower in height. Thus, it is hard to tell the height difference between the ground and the crops. In such cases, the color-based segmentation will be the only solution.

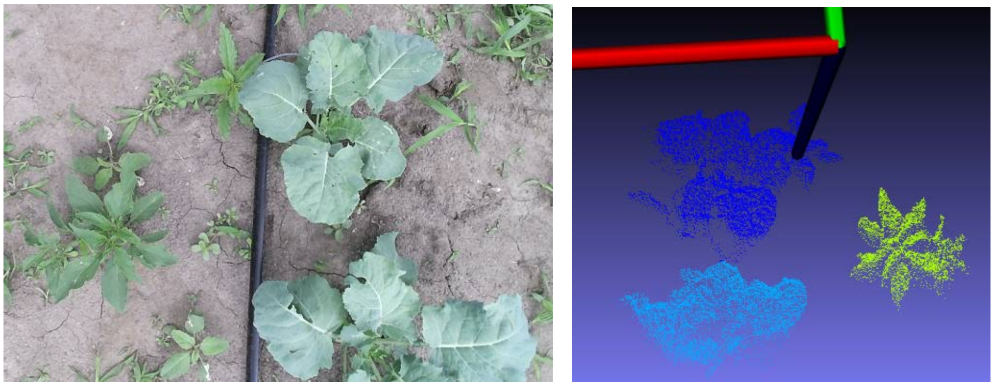


Figure 10. Sample segmentation successes. The left figure shows the one color image, and the right figure shows the segmentation result point cloud. The background is eliminated, and the pixels belonging to crops are extracted and separated as clusters.

Figure 11. Sample segmentation failures. The left figure shows a segmentation failure case in tomato field, in which the ground is covered with weeds. The right figure shows a lettuce field case, in which the depth-based segmentation is failed, because of the low crops height.

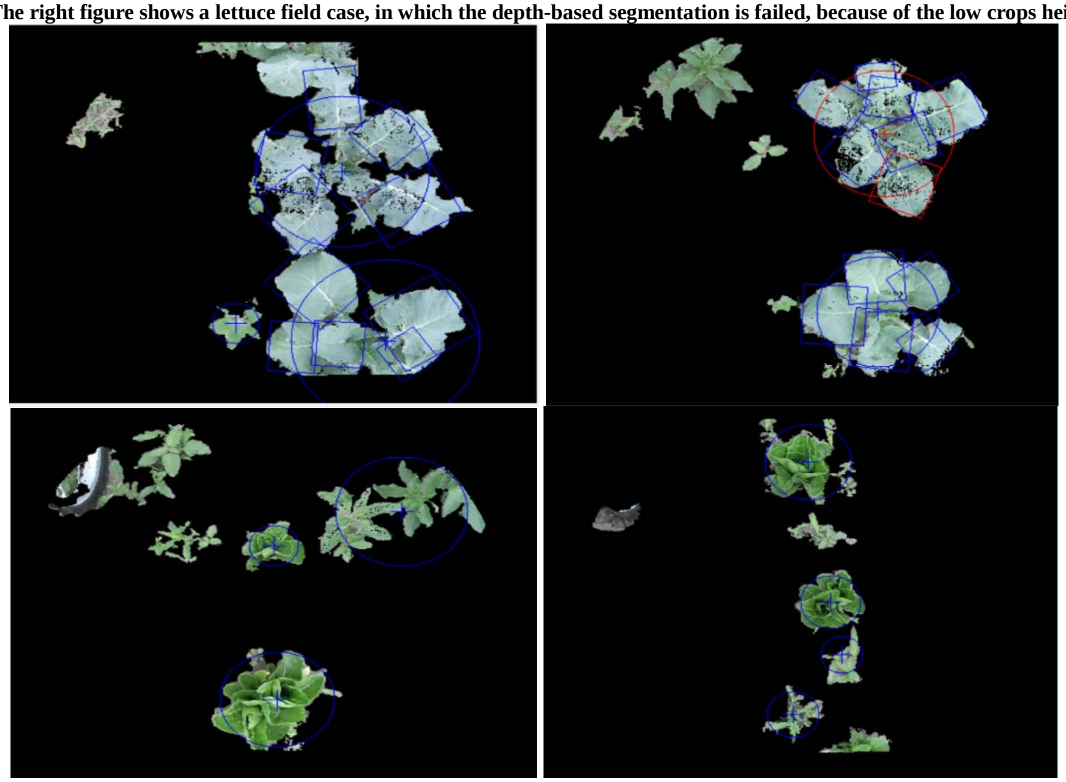


Figure 12. Examples of the candidate plants detection and localization run-time result. In those images, the detected plants are circled the positions are labeled with crosses. For detecting Broccoli, segmented leaves are also labeled with squares.

Crop plants classification and localization

Using the data collected, crop plant classification and localization systems are developed for two crop species: broccoli and lettuce (Figure 12). After testing the developed algorithms with the collected data, 92.5% of actual broccoli and lettuce crop plants were detected, and locating error of 15.9 mm and 8.5 mm were obtained for broccoli and lettuce, respectively (Table 4).

There are several error sources contributed to the miss-detection errors and locating errors. The first source is error from the Kinect V2 sensor. Since working outdoors will increase the noise level of Kinect V2, even with sunlight shaded; The second is error from the algorithms. With the disturbances from weeds, sometimes the image processing algorithm will fail in detecting ground and extracting leaves. The third is that the ground is not strictly flat in the field, which will also affect the ground fitting results.

Table 4. The detection rate and average localization error for broccoli and lettuce dataset at different growth stages.

Collection data	Days after transplant	Broccoli Detection rate/ localization error	Lettuce Detection rate/ localization error
June 13, 2016	12	100% / 14.8 mm	100% / 5.2 mm
June 17, 2016	16	98% / 10.6 mm	96% / 6.9 mm
June 23, 2016	22	87% / 17.2 mm	94% / 9.9 mm
June 27, 2016	26	85% / 21.1 mm	80% / 12.1 mm

During crop/weed classification, 403 sets of plant features were extracted from the broccoli field dataset, with 289 of them are supervised as broccoli crops, and others are supervised as non-crops. 861 sets of plant features were extracted from the lettuce field dataset, with 376 sets of them supervised as lettuce crops, and the others supervised as non-crops. The classifications for different species are performed separately. Seven classification models are applied in this study. They are logistic regression (LR), k-nearest neighbor (KNN), artificial neural network (ANN), Bayes classifiers QDA, support vector machine (SVM), random forest and tree-based AdaBoost. After fitting models with different methods and tuned to minimize the CV result, the best performance of each model is listed in tables.

Broccoli classification

The broccoli classification is separated into two stages in this study, one is leaf classification, which uses leaf features only, and the other is plant classification, which combines all the information.

Table 5. Models evaluation and comparison result for Broccoli leaf classification. Listing tuning parameters, training errors and CV errors. Adaboost model yields the best performance.

Model	Tuning parameter	Training error	CV error
LR	None	0.076	0.115
ANN	Layer = (5, 2)	0.060	0.192
KNN	K=5	0.110	0.134
SVM	Kernel = radial Gamma = 0.01 Cost = 100	0.062	0.135
QDA	none	0.081	0.127
RandomForest	N=500	0.152	0.199
AdaBoost	TreeDepth=4 Iter = 100 Nu = 0.2	0.084	0.106

As a result (Table 5), AdaBoost performs the best in leaf classification. After classifying each leaf, for each plant, the classified crop leaf parameters will form a new set of predictors for plant classification. The color features are not stable for leaves acording to out test, thus they are not used for plant classification.

In plant classification, some of the predictors came from the result of leaf classification. But for some plants, the leaves are failed to be segmented. Then the plant classification becomes a problem with missing data, for which only some tree-based classification methods (Decision tree and AdaBoost) are compatible. In this study, LR, KNN, ANN, SVM, QDA, RandomForest are tested without leaf parameters, and AdaBoost is tested with leaf parameters.

As a result (Table 6), AdaBoost performs the best in Broccoli plant classification and achieved an error rate of 5.5% in average during Cross-Validation.

Table 6. Models evaluation and comparison result for Broccoli plant classification. Listing tuning parameters, training errors and CV errors.

Model	Tuning parameter	Training error	CV error
LR	None	0.092	0.123
ANN	Layer = (5, 2)	0.051	0.131
KNN	K=3	0.125	0.163
SVM	Kernel = radial Gamma = 0.01 Cost = 100	0.056	0.076
QDA	none	0.064	0.124
RandomForest	N=500	0.097	0.104
AdaBoost	TreeDepth=4 Iter = 50 Nu = 0.1	0.04	0.055

Lettuce classification

Since only a few leaves are extracted from Lettuce dataset, the plant classification is performed without leaf features. As a result, AdaBoost performed the best in Lettuce plant classification and achieved an error rate of 6.8% in average during Cross-Validation.

Table 3-1. Models evaluation and comparison result for lettuceplant classification. Listing tuning parameters, training errors and CV errors. AdaBoost performed the best in plant classification.

Model	Tuning parameter	Training error	CV error
LR	None	0.112	0.134
ANN	Layer = (5, 2)	0.079	0.173
KNN	K=5	0.105	0.135
SVM	Kernel = radial Gamma = 0.01 Cost = 100	0.087	0.154
QDA	none	0.072	0.115
RandomForest	N=500	0.097	0.127
AdaBoost	TreeDepth=4 Iter = 50 Nu = 0.1	0.051	0.068

The reasons why AdaBoost performed the best are: Firstly, crop/non-crop classification is a two-class classification problem, with non-linear decision boundary. Crops are in the same patterns, and others are all non-crops. Then using liner models (LR, QDA, ADNN, SVM) with sigmoid-shaped responses is not suitable. And Adaboost is designed to create complex decision boundaries. Secondly, in this study, only about ten predictors are used, with limited information available. Thus, Random Forest, which is mainly used with many predictors available and to find the most valuable features, is less suitable than AdaBoost.

The classification error comes from four sources: The first is sensor error. As the data is collected outdoors with Kinect V2 sensor, the noise can affect the accuracy even with sunlight shaded. Also, the resolution of depth image is still limited. The second is from the image processing algorithms. As the situations are complex in the field, the algorithms such as ground fitting and leaf extraction may have inaccurate results, then affect the accuracy of extracted features. The last source is that insufficient features are extracted. The features extracted in this study still cannot fully parameterize the crops and non-crops including all differences.

Conclusion

In this study, a computer vision system was developed based on Kinect V2 sensor, using the fusion of two-dimensional textural data and three-dimensional spatial data to recognize and localized crop plants different growth stages. Thousands of pictures were taken using the Kinect v2 sensor in this reporting period. The images were taken at different growth stages, from germination or transplantation to maturity. Using the data collected, solutions are developed for some of the crop species in discrimination and localization. A classification system was developed using supervised machine learning technique, with features of plant color and morphology as input, to realize discrimination of crop plants at different growth stages. Different classification methods are also evaluated and compared in this application. The results show that this plant discrimination system is feasible for application in robotic weeding. After testing the algorithm with the collected data, 92.5% of actual broccoli and lettuce crop plants were detected, and locating error of 15.9 mm and 8.5 mm were obtained for broccoli and lettuce, respectively. By evaluating and comparing different classification methods, AdaBoost was found the one who performed best in this application. After feature extraction and classification, crops can be recognized with an error rate of 5.5% and 6.8% with AdaBoost for Broccoli and Lettuce in average, respectively.

Acknowledgements

The research is sponsored by Leopold Center for Sustainable Agriculture, Grants No. M2009-23, M2012-24. USDA NIFA Foundational Programs, Grant No. 20136702121126

References

Alenyà, G., Dellen, B., Foix, S., & Torras, C. (2012). *Robotic Leaf Probing Via Segmentation of Range Data Into Surface Patches.* Paper presented at the IROS Workshop on Agricultural Robotics: Enabling Safe, Efficient, Affordable Robots for Food Production, Vilamoura, Portugal.

Blatt, M., Wiseman, S., & Domany, E. (1996). Superparamagnetic Clustering of Data. *Physical Review Letters, 76*(18), 3251-3254.

Bradski, G., & Kaehler, A. (2008). *Learning OpenCV: Computer vision with the OpenCV library*: " O'Reilly Media, Inc.".

Corti, A., Giancola, S., Mainetti, G., & Sala, R. (2015). A metrological characterization of the Kinect V2 time-of-flight camera. *Robotics and Autonomous Systems*, Available online

Du, J., Wang, X., & Zhang, G. (2007). Leaf shape based plant species recognition. *Applied Mathematics and Computation*, 883-893.

Fankhauser, P., Bloesch, M., Rodriguez, D., Kaestner, R., Hutter, M., & Siegwart, R. (2015). *Kinect v2 for mobile robot navigation: Evaluation and modeling.* Paper presented at the Advanced Robotics (ICAR), 2015 International Conference on.

Finlayson, G. D., Hordley, S. D., & Drew, M. S. (2002). *Removing Shadows from Images*. Paper presented at the Computer Vision — ECCV 2002.

Jin, J., & Tang, L. (2009). Corn Plant Sensing Using Real-Time Stereo Vision. *Journal of Field Robotics*, 591–608.

Kazmi, W., Foix, S., Alenyà, G., & Andersen, H. J. (2014). ndoor and outdoor depth imaging of leaves with time-of-flight and stereo vision sensors: Analysis and comparison. *ISPRS Journal of Photogrammetry and Remote Sensing*, 128-146.

Li, J. (2014). *3D machine vision system for robotic weeding and plant phenotyping*. Retrieved from

Li, L. (2014). Time-of-Flight Camera – An Introduction. *Technical White Paper*.

McErlich, A. F., & Boydston, R. A. (2014). Current state of weed management in organic and conventional cropping systems *Automation: The Future of Weed Control in Cropping Systems* (pp. 11-32): Springer.

Nakarmi, A., & Tang, L. (2012). Automatic Inter-plant spacing sensing at early growth stages using a 3D vision sensor. *Computers and Electronics in Agriculture*.

Nielsen, M., Andersen, H. J., Slaughter, D. C., & Giles, D. K. (2004). *Detecting leaf features for automatic weed control using trinocular stereo vision.* Paper presented at the Precision Agriculture and Other Precision Resources Management, Minneapolis, MN, USA.

Philipp, I., & Rath, T. (2002). Improving plant discrimination in image processing by use of different colour space transformations. *Computers and Electronics in Agriculture, 35*(1), 1-15. doi:http://dx.doi.org/10.1016/S0168-1699(02)00050-9

Rusu, R. B. (2009). *Semantic 3D Object Maps for Everyday Manipulation in Human Living Environments*. Retrieved from

Sansoni, G., Trebeschi, M., & Docchio, F. (2009). State-of-The-Art and Applications of 3D Imaging Sensors in Industry, Cultural Heritage, Medicine, and Criminal Investigation. *Sensors*, 9, 568-601.

Scotford, I. M., & Miller, P. C. H. (2005). Applications of Spectral Reflectance Techniques in Northern European Cereal Production: A Review. *Biosystems Engineering*, 235-250.

Slaughter, D. C. (2008). Autonomous robotic weed control systems: A review. *computers and electronics in agriculture*.

Sui, R., Thomasson, J. A., Hanks, J., & Wooten, J. (2008). Ground-based sensing system for weed mapping in cotton. *Computers and Electronics in Agriculture*, 31–38.

Surmann, H., & Nüchter, A. (2003). An autonomous mobile robot with a 3D laser range finder for 3D exploration and digitalization of indoor environments. *Robotics and Autonomous Systems*, Volume 45, Issues 43–44.

USDA-NASS (Producer). (2014). http://www.nass.usda.gov/Publications/Ag_Statistics/2014/. *2014 Agricultural Statistics Annual*.

Weiss, U., & Biber, P. (2011). Plant detection and mapping for agricultural robots using a 3D LIDAR sensor. *Robotics and Autonomous Systems*, 265-273.

Zimdahl, R. L. (2013). *Fundamentals of Weed Science*: Elsevier Science.

Zwiggelaar, R. (1998). A review of spectral properties of plants and their potential use for crop/weed discrimination in row-crops. *Crop Protection*, 189-206.